DOCUMENT RESUME

ED 174 675 TM 009 647

AUTHOR Newman, Isadore; And Others
TITLE A Monte Carlo Evaluation of Estimated Parameters of Five Shrinkage Estimate Formuli.
PUB DATE Apr 79
NOTE 21p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Correlation; *Mathematical Formulas; *Multiple Regression Analysis; Prediction; Sampling; Simulation; *Statistical Bias
IDENTIFIERS *Cross Validation; Monte Carlo Method

ABSTRACT

A Monte Carlo study was conducted to estimate the efficiency of and the relationship between five equations and the use of cross validation as methods for estimating shrinkage in multiple correlations. Two of the methods were intended to estimate shrinkage to population values and the other methods were intended to estimate shrinkage from sample to sample. Comparisons were made for the six methods with three levels of population multiple correlations, four different sample sizes, and three different degrees of multicollinearity (average correlations among the predictor variables). One hundred samples were drawn for each of the 48 conditions. The data showed no advantage for the cross-validation procedure in comparison to the other methods. The major purpose of the study was to check the differential effects of multicollinearity on shrinkage estimates; but it was found to have virtually no effect. This may have been an artifact of the method used to generate the data. Caution is recommended against over-generalization from artificially generated data. (The computer program is appended). (CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED174675

A MONTE CARLO EVALUATION OF ESTIMATED PARAMETERS OF FIVE SHRINKAGE ESTIMATE FORMULI

ISADORE NEWMAN - UNIVERSITY OF AKRON
KEITH A. MCNEIL - NATIONAL TESTING SERVICE
THOMAS GARVER - FIRESTONE PERSONNEL AND TRAINING DIVISION
GAYLE A. SFYMOUR - UNIVERSITY OF AKRON COMPUTER CENTER

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Isadore Newman
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ABSTRACT

This study employs a Monte Carlo simulation to determine the accuracy with which the shrinkage in R^2 can be estimated by five shrinkage formuli and cross-validation. The study dealt with the use of shrinkage and cross-validation for different sample sizes, different R^2 values, and different degrees of multi-collinearity.

The present study was designed to determine the relative effects of four different levels of multicollinearity and population R^2's on six different shrinkage estimates. Two of these estimates are used predominately to estimate the population R^2 from the sample (McNemar (M) 1949,1962; and Wherry/McNemar (W/M) 1931). The four others are traditionally used to estimate shrinkage from sample to sample (Lord1 (L_1); Lord2 (L_2), Darlington (D) and Cross-Validation (CV_2). In addition this paper will look at how well the sample to sample shrinkage estimates predict the true population R^2.

TM009 647

This paper was presented at the annual meeting of the American Educational Research Association; April 1979, San Francisco, Ca.

Session 16.30

It was assumed that cross-validation would be the least effected by multicollinearity, however, the study did not support this hypothesis. Apparently, multicollinearity had very little effect on each of the shrinkage estimates.

Applied researchers who use multiple regression frequently are interested in:

1) estimating the relationship between a set of predictive variables to a criterion in a sample for the purpose of generalizing to the population

2) estimating the magnitude of relationships from the sample to the population

3) estimating accuracy of prediction from sample to sample and then to the population

In both estimating the magnitude of relationships and the accuracy of prediction, shrinkage estimates are both utilitarian and potentially misleading. Since there are a variety of shrinkage estimating procedures and formuli it is important to understand the differential effects of the shrinkage estimates for different situations and try to identify which is most appropriate under different conditions.

In the classical article by Charles Mosier (1951) he defines six purposes for doing cross validations. Similarly he defines six different types of cross validations.

1) Cross validation using weights determined in one sample applied in a second sample drawn from the same population to determine the differences in R^2 between the two samples (this type of shrinkage is the one that will be used in this study)

2) Validity generalization is similar to the first but the samples come from two distinctly different populations

3) Validity extention in which the samples come from different populations and the two samples criteria are also different

4) Simultaneous validation in which the purpose is to select a set of items (batteries) effectiveness in two different samples from different populations for the purpose of having the single best most useful battery for both populations

5) Replication is a method of determining the best set of weights by drawing a variety of samples from the same population and determining the best combination of weights by using information from these several samples

6) Double-cross validation in which two samples are drawn simultaneously and the weights calculated for each sample are then applied to the other sample

The major problem with cross-validation is that when one estimates shrinkage one tends to estimate the weights upon half of the available data when it is well known that the stability of the weight increases as the ratio of the number of subjects increases in relationship to the number of variables (producing less shrinkage). A suggested path around this dilemma has been to use double-cross-validation. If the weights in each sample are applied to the other sample (weights of Sample 1 to Sample 2 and vice versa) without too much shrinkage (subjective decision) then one may combine the samples and calculate the final weights on the total sample. Another suggestion is to look at which weights (Sample 1 when applied to Sample 2 or Sample 2 when applied to Sample 1) produces the least amount of shrinkage and then use those weights.

The assumption made by Mosier is that the best way to estimate

the accuracy of the weights in one sample to predict the other sample is by using separate samples. He bases this argument on the work done by Cureton (1950) and Kurtz (1948). This assumption is questioned by Dalington (1969), Lord (1950), Schmitt, Coyle and Rauschenberger (1977), and Claudy (1978). If a mathematical correction can be used based upon the entire sample that can predict the shrinkage in R^2 from sample to sample accurately, it would be preferable to use cross-validation procedure which requires two samples. It would also produce more stable weights since it would be done on the entire sample instead of just half the sample. This mathematical correction procedure has been criticized because of some simple misunderstandings.

As stated by others (Newman, 1973; Uhl, 1970, Lord, 1950, Darlington, 1968, etc.), there are a variety of mathematical shrinkage estimate corrections, which we shall call McNemar, Wherry-McNemar, Lord 1, Darlington, and Lord 2.

$$\hat{R}^2 = 1 - (1-R^2)\frac{N-1}{N-K} \quad \text{(McNemar 1949)}$$

$$\hat{R}^2 = 1 - (1-R^2)\frac{N-1}{N-K-1} \quad \text{(Wherry/McNemar (W/M) 1931)}$$

$$\hat{R}^2 = 1 - (1-R^2)\frac{N+K+1}{N-K-1} \quad \text{(Lord 1 } (L_1)\text{)}$$

where:

R^2 = the corrected estimate of the multiple correlation.

R = the actual calculated multiple correlation.

K = the number of independent variables.

N = the number of linearly independent observations.

Also, Schmitt and Rauschenberger (1977) presented several shrinkage estimation formuli which will be included in the analysis. These were:

$$\rho c^2 = 1 - \frac{N-1}{N-p-1}\ \frac{N-2}{N-p-2} \times \frac{N+1}{N}\ (1-R^2) \qquad \text{(Darlington (D))}$$

$$\rho c^2 = 1 - \frac{N-1}{N-p-1}\ \frac{N+p+1}{N}\ (1-R^2) \qquad \text{(Lord 2 } (L_2) \text{)}$$

where:

ρc^2 = the correlated estimate of the multiple correlation.

R = the actual calculated multiple correlation.

N = the number of linearly independent observations.

P = the number of independent variables.

The McNemar and Wherry-McNemar's shrinkage estimate was developed for and is only appropriate to use when one is interested in predicting a population value from a sample. The others, Lord 1, Lord 2, and Darlington were derived for the purpose of estimating shrinkage from sample to sample, that is the stability and accuracy of the weights from one sample predicting to another.

A Monte Carlo study was run to estimate the efficiency and relationship between these five equations and cross-validation in predicting known population values for differing population R^2's, differing levels of correlations between four predictor variables (degree of multicollinearity) and for four different N sizes (14, 30, 50, and 100). To summarize, the Monte Carlo design will look at six estimates of shrinkage (McNemar, Wherry-McNemar, Lord 1, Lord 2, Darlington, and Cross-Validation) for three different ranges of know population R^2's (Situation 1's R^2's = .06, .07, .06,

.08; Situation 2 R^2's = .31, .32, .33, .34, and Situation 3 R^2's = .45, .47, .46, and .55). For each of the three Situations, there were four differing correlations between the four predictor variables (estimates of multicollinearity ($\bar{r}$)). These were for Situation 2, the correlations ($\bar{r}$) between the four predictor variables were .13, .30, .52, .80; and for Situation 3, the correlations ($\bar{r}$) were .15, .29, .53, .82. For each of the twelve populations (that is a particular population R^2 for a particular level of multicollinearity) there are four different N sizes (14, 30, 50, and 100). Therefore, the Monte Carlo Study is a 3x4x4 design, a range of three sets of population values, a range of four different multicollinearity conditions for four different N sizes. (see Table #1 for results) The data is Table #1 is based upon 100 random samples picked for each situation from each of its 12 specific population. The r for multicollinearity is the average for the 100 samples. In total there were 4,800 samples drawn (3x4x4x100 (samples)).

TABLE OF SHRINKAGE ESTIMATES

Situation I

Population 1		N	R^2_S	R^2_M	$R^2_{W/M}$	$R^2_{L_1}$	R^2_D	$R^2_{L_2}$	R^2_{CV1}	R^2_{CV2}
R^2 = .06	A	14	.33	.13	.04	-.41	-.43	-.31	.69	.15
	B	30	.17	.08	.04	-.16	-.13	-.12	.32	.08
$\bar{r}$ = .03	C	50	.14	.08	.06	-.05	-.05	-.03	.22	.06
	D	100	.09	.06	.05	-.00	.00	.01	.13	.04
Population 2										
R^2 = .07	A	14	.37	.19	.10	-.32	-.50	-.23	.68	.17
	B	30	.19	.10	.06	-.14	-.14	-.10	.32	.09
$\bar{r}$ = .29	C	50	.13	.08	.05	-.06	-.06	-.04	.21	.07
	D	100	.10	.07	.06	.00	.00	.01	.13	.05
Population 3										
R^2 = .06	A	14	.34	.15	.05	-.38	-.48	-.29	.70	.17
	B	30	.19	.10	.06	-.13	-.13	-.09	.31	.08
$\bar{r}$ = .54	C	50	.13	.08	.05	-.06	-.06	-.04	.22	.06
	D	100	.10	.07	.06	.01	.01	.02	.13	.05
Population 4										
R^2 = .08	A	14	.35	.16	.07	-.36	-.51	-.27	.69	.20
	B	30	.20	.11	.07	-.12	-.12	-.08	.36	.08
$\bar{r}$ = .82	C	50	.13	.07	.05	-.06	-.06	-.04	.23	.06
	D	100	.11	.08	.07	.02	.02	.03	.15	.06

Where: R^2_S = The R^2 for the sample

R^2_M = McNemar's shrinkage estimate

$R^2_{W/M}$ = Wherry/McNemar's shrinkage estimate

$R^2_{L_1}$ = Lord 1 shrinkage estimate

continued on next page

Situation II

Population 5		N	R^2_S	R^2_M	$R^2_{W/M}$	$R^2_{L_1}$	R^2_D	$R^2_{L_2}$	R^2_{CV1}	R^2_{CV2}
R^2 = .34	A	14	.50	.35	.28	-.05	-.17	.02	.76	.21
	B	30	.40	.33	.30	.16	.16	.18	.48	.24
$\bar{r}$ = .13	C	50	.39	.35	.34	.26	.26	.27	.43	.29
	D	100	.37	.35	.34	.30	.30	.31	.40	.30

Population 6										
R^2 = .32	A	14	.51	.36	.29	-.04	-.15	.04	.74	.26
	B	30	.41	.34	.32	.18	.18	.20	.51	.21
$\bar{r}$ = .30	C	50	.35	.31	.29	.20	.20	.22	.42	.24
	D	100	.34	.32	.31	.27	.27	.28	.37	.30

Population 7										
R^2 = .31	A	14	.50	.35	.28	-.05	-.12	.02	.76	.23
	B	30	.38	.31	.28	.13	.13	.16	.45	.26
$\bar{r}$ = .32	C	50	.34	.30	.29	.20	.20	.21	.39	.24
	D	100	.33	.31	.30	.26	.26	.26	.36	.27

Population 8										
R^2 = .33	A	14	.50	.34	.27	-.07	-.15	.01	.71	.25
	B	30	.41	.34	.31	.17	.17	.20	.48	.24
$\bar{r}$ = .80	C	50	.36	.32	.30	.22	.22	.23	.42	.25
	D	100	.36	.34	.33	.29	.29	.30	.37	.30

Continued from Table I:

R^2_D = Darlington's shrinkage estimates

R^2_{CV1} = R^2 calculated on 1/2 of the sample

R^2_{CV2} = The cross validated shrinkage estimate based on the weights from the first sample (CV1) applied to the second sample

Situation III

Population 9		N	R^2_S	R^2_M	$R^2_{W/M}$	$R^2_{L_1}$	R^2_D	$R^2_{L_2}$	R^2_{CV1}	R^2_{CV2}
R^2 = .45	A	14	.59	.47	.41	.13	.05	.20	.80	.33
	B	30	.51	.46	.43	.32	.31	.34	.55	.36
$\bar{r}$ = .15	C	50	.48	.45	.44	.37	.37	.38	.52	.39
	D	100	.46	.45	.44	.41	.41	.41	.46	.41

Population 10										
R^2 = .46	A	14	.61	.50	.44	.18	.08	.24	.80	.29
	B	30	.50	.44	.42	.29	.29	.32	.53	.37
$\bar{r}$ = .29	C	50	.49	.46	.45	.38	.38	.39	.53	.39
	D	100	.47	.46	.46	.42	.42	.43	.49	.45

Population 11										
R^2 = .47	A	14	.62	.50	.45	.19	.10	.25	.81	.34
	B	30	.54	.48	.46	.35	.35	.37	.59	.40
$\bar{r}$ = .53	C	50	.51	.48	.46	.40	.40	.41	.53	.42
	D	100	.49	.48	.47	.44	.44	.45	.50	.45

Population 12										
R^2 = .55	A	14	.66	.56	.52	.29	.20	.34	.84	.43
	B	30	.58	.53	.51	.41	.41	.43	.63	.44
$\bar{r}$ = .82	C	50	.57	.54	.53	.47	.47	.48	.60	.51
	D	100	.55	.54	.53	.51	.51	.51	.56	.53

Table 1 seems to indicate that the McNemar and Wherry/McNemar tend to be relatively stable for the four different N sizes in Situation 1, with the Wherry/McNemar being more consistent across all four N's when the R^2 (population) is close to zero (.06). For Situation 2 the R^2's ranges from .31 to .34 and in Situation 3 the R^2's range from .45 to .55. In both Situations 2 and 3 the relationship of the McNemar and Wherry/McNemar to the population R^2's seem to become even more similar than they were in Situation 1. Lord 1, Lord 2 and Darlington, which are generally used for sample to sample estimates instead of sample to population, tend to have higher R^2's as the N's increase from 14 to 100. However, for all four populations in Situation 1, these are underestimates which one would expect since they are estimates from sample to sample and not sample to population. For Situations 2 and 3 when the N's were 50 and 100, Lord1 and Darlington tended to be fairly decent estimates of the population R^2, but they underestimated the population more than does the McNemar and the Wherry/McNemar.

Surprisingly to us, the cross-validation shrinkage estimate (CV_2) for Situation 1 with an N=14 tends to overestimate the population R^2. For the N=30, N = 50, and N=100, the CV_2 is fairly accurate. For Situations 2 and 3, CV_2 are underestimates of the population R^2's for all N sizes. An interesting effect is that, as the N's become larger, the CV_2 values become smaller, unlike Lord 1, Lord 2, and Darlington, whose values become larger as the N's increase (for Situation 1 only). It seems that the CV_2 values, except for Situation 1, tend to be more consistent than the mathematical values across all N sizes.

However, the shrinkage (amount of shrinkage) between CV_1 and CV_2 tend to decrease as the N's get larger.

The degree of multicollinearity ($\bar{r}$), that is the correlation between predictor variables, seems to have virtually no detectable effect on the accuracy of any of the shrinkage estimates.

Brief Discussion

Assuming that the data have been generated correctly (see the appendix for the computer program) the data from Table 1 would then strongly force the authors to question the preference for using cross-validation as a standard since.

1) The shrunken R does not seem to be more accurate than the mathematical approaches (Darlington (1969), Schmitt and Coyle (1977), and Claudy (1978)).

2) The shrinkage does not decrease as much proportionally as the authors expected when the N's increase

3) Cross-validation forces one to split the sample in half which tends to produce less stability than one would get using the entire sample.

Suggestions for Further Research

Darlington (1978) discusses Stein-type regression which is a shrinkage corrected regression on the weights instead of on the R^2. The partial regression weights are multiplied by 1 - S.

where: S is a correction for shrinkage obtained from the following formula

$$S = \frac{p-2}{n+2-p} \times \frac{(1-R^2)}{R^2}$$

where: p = number of predictor variables

n = size of the sample

R = sample multiple R (not corrected)

This procedure reduces the variability of the weights from sample to sample. As stated previously, in this procedure, all of the weights for the predictor variables are multiplied by the same constant (1-S). The purpose of Stein regression is identical to that of Ridge regression (Newman and Fraas (1979) , Walton, Newman and Fraas (1978); Hoerl and Kennard (1970)) except that in Ridge regression all of the weights are not multiplied by the same constant. Instead they are multiplied by a factor that is determined by the importance of each variable's contribution to the multiple R. It seems that Stein regressions simplicity and ease in determining a constant can be very valuable. The authors are suggesting that one may be interested in investigating the most appropriate shrinkage estimates for the weights, given different conditions, as has been and is being done in investigating the most appropriate shrinkage for the multiple R, given differing constraints.

As with most Monte Carlo studies, one should be careful about not overly generalizing from the artificially generated data to the real world data. It is possible that there is something idiosyncratic in the programs production of the data that one has to at least consider. The data should be replicated to see how well the shrinkage estimates predict from sample to sample.

The major purpose of this study was to check the differential effects of multicollinearity on shrinkage estimates. We found that multicollinearity has virtually no effect on shrinkage estimates. This may be an artifact of the data generated and this should be considered. In addition, a variety of other shrinkage estimates

which were not considered in this paper may be of value for future studies. Such estimates may include:

Burket's (1964) - weight validity estimate:

$$\rho_c^2 = \frac{Nr^2-K}{r(N-K)}$$

Olkin and Pratt (1958):

$$\rho^2 = 1-\frac{N-3}{N-K-1}\ (1-r^2)\ F\ (1,1;\ \frac{N-K+1}{2};\ 1-r^2)$$

Pratt's (1964) approximation of the Olkin and Pratt unbiased estimate of ρ^2:

$$\rho^2 = 1 - \frac{(N-3)(1-r^2)}{N-K-1}\left[1+\frac{2(1-r^2)}{N-K-2.3}\right]$$

Herzberg (1969) which is a further approximation of the Olkin and Pratt:

$$\rho^2 = 1 - \frac{(N-3)(1-r^2)}{N-K-1}\left[1+\frac{2(1-r^2)}{N-K+1}\right]$$

It is hoped that this paper has at least raised some additional questions that may lead to fruitful further investigations.

References

Burket, G.R. A study of reduced rank models for multiple prediction. Psychometric Monograph, 1964, 12.

Claudy, J.G. Multiple regression and validity estimation in one sample. Applied Psychological Measurement, 1978, vol 2, 4 595-607.

Darlington, R.B. Reduced-variance regression. Psychological Bulletin, 1978, vol 85, 6.

Herzberg, P.A. The parameters of cross-validation. Psychometric Monograph, 1969, 16.

Hoerland and Kennard Ridge Regression: applications to nonorthogonal problems. Technometrics, 1970, vol 12, 1.

Klein, M. and Newman, I. Estimated parameters of three shrinkage estimate formuli. Multiple Linear Regression Viewpoints, 1974, 4, 6-11.

Lord, F.M. Efficiency of prediction when a regression equation from one sample is used, in a new sample. Research Bulletin, 1950, vol 50, 40, Princeton, N.J.: E.T.S.

McNemar, A. Psychological Statistics (3rd edition) New York: Wiley and Sons 1962.

Mosier, C.I. Problems and designs of cross-validation. Educational and Psychological Measurement, 1951, vol 11, 1-11.

Newman, I. Variations between shrinkage estimation formula and the appropriateness of their interpretation. Multiple Linear Regression Viewpoints, 1973, 4, 45-48.

Newman, I. and Fraas, J.W. Some applied research concerns using multiple linear regression analysis. Paper presented at Psychology Division of The Ohio Academy of Science, Capital University, Columbus, Ohio, April 1977.

Olkin, I. and Pratt, J.W. Unbiased estimation of certain correlation coefficients. Annuals of Mathematical Statistics, 1958, vol 29, 201-211.

Pohlman, J.T., and Moore, J.F. Interval estimation of the population squared multiple correlation. Multiple Linear Regression Viewpoints, 1977, 8, 18-24.

Schmitt, N., Coyle, B.W., and Rauchenberger, J. A Monte Carlo evaluation of three formula estimates of cross-validated multiple correlation. Psychological Bulletin, 1977, 84, 751-758.

Uhl, N. and Eisenber, T. Predicting shrinkage in the multiple correlation coefficient. Educational and Psychological Measurements 1970, vol 30.

Walton, J.M., Newman, I. and Fraas, J. Ridge Regression· A Panacea? M.L.R.V. 1978, vol 8, 2, 118-15.

```
      ∇GAUSS[□]∇
      ∇ GAUSS X;M;S;MIN;MAX;POP1;POP;N
[1]   M←X[1]
[2]   S←X[2]
[3]   MIN←X[3]
[4]   MAX←X[4]
[5]   POP←X[5]
[6]   POP1←POP
[7]   V←⍳0
[8]  TRYAGAIN:V←V,⌊0.5+(M+(S×((+/(0.1×((?(POP1,12)⍴11)-1)))-6)))
[9]   N←⍴V←((V<MAX)∧(V>MIN))/V
[10]  POP1←(POP-N)
[11]  →(POP1>0)/TRYAGAIN
[12]  'A POPULATION OF ',(▼POP),' MEMBERS HAS BEEN CREATED AND IS STORED IN THE VARIABLE V'
      ∇

      ∇CREATE[□]∇
      ∇ A CREATE B;K
[1]   K←1
[2]   MAT←((1↑⍴A),5)⍴0
[3]  LOOP:K←K+1
[4]   B[1] RAND 1↑⍴A
[5]   MAT[;K]←B+A
[6]   →LOOP×⍳K<5
[7]   B[2] RAND 1↑⍴A
[8]   MAT[;1]←A+B
[9]   10 4 ▼R←CM MAT
[10]  R←(¯4++/+/R[2 3 4 5 ; 2 3 4 5])÷12
[11]  'AVERAGE INTER-CORRELATION (R)= ', 6 4 ▼R
[12]  T← 5 4 3 2 1 REG MAT
[13]  'MULTIPLE R SQ = ', 7 5 ▼T[8;5]
      ∇
```

```
      ∇RAND[⎕]∇
      ∇ NN RAND N;K
[1]   K←0
[2]   R←,⍳0
[3]   →NUL×⍳NN=0
[4]   LOOP:K←K+1
[5]   R←R,?NN
[6]   →LOOP×⍳K<N
[7]   R←R-NN×2
[8]   →0
[9]   NUL:R←N⍴0
      ∇


      ∇REG[⎕]∇
      ∇ T←V REG X;N;U;M;R;Q;S;B
[1]   T←((3+⍴V),5)⍴0
[2]   T[(⍳⍴V);1]←V[⍋V],¯1↓V
[3]   M←(+⌿U←X[;V])↓N←(⍴X)[1]
[4]   Q←(⍳N)∘.×R←U-(⍴U)⍴M
[5]   T[(3+⍴V);3]←Q[(⍴V);⍋V]
[6]   S←⍉0 ¯1↓Q←¯1 0↓Q
[7]   B←(-M+.×B,¯1),B←,S+.×Q[;⍋V]
[8]   T[(⍳⍴V);2]←B
[9]   T[(1+⍴V);3]←(1↓B)+.×Q[;⍋V]
[10]  T[((⍴V)+⍳3);2]←((⍴V)-1),(N-⍴V),N-1
[11]  T[(2+⍴V);3]←-/T[((3+⍴V),1+⍴V);3]
[12]  T[((⍴V)+⍳2);4]←T[((⍴V)+⍳2);3]↓T[((⍴V)+⍳2);2]
[13]  T[(1+⍳(⍴V)-1);3]←(×B)×(|B←T[(2+⍴V);4])×+/S×(⍳(⍴V)-1)∘.=⍳(⍴V)-1)*0.5
[14]  T[(1+⍳(⍴V)-1);4]←T[(1+⍳(⍴V)-1);2]↓T[(1+⍳(⍴V)-1);3]
[15]  T[(3+⍴V);4]←(×T[(2+⍴V);4])×(|T[(2+⍴V);4])*0.5
[16]  T[(1+⍴V);5]←T[(1+⍴V);4]↓T[(2+⍴V);4]
[17]  T[(3+⍴V);5]←T[(1+⍴V);3]↓T[(3+⍴V);3]
      ∇
```

```
      ∇SHRINK[□]∇
      ∇ POP SHRINK N
[1]   V←8ρ0
[2]   I←0
[3]   N2←⌈N[1]÷2
[4]   GO:NN←RAND1 N[1]
[5]   SAMP←POP[NN;]
[6]   RS← 5 4 3 2 1 REG SAMP
[7]   V[1]←V[1]+RS[8;5]
[8]   V[2]←V[2]+1-(1-RS[8;5])×(N[1]-1)×÷(N[1]-4)
[9]   V[3]←V[3]+1-(1-RS[8;5])×(N[1]-1)×÷(N[1]-3)
[10]  V[4]←V[4]+1-(1-RS[8;5])×(N[1]+5)×÷(N[1]-3)
[11]  V[5]←V[5]+1-((N[1]-1)×÷(N[1]-3))×((N[1]-2)×÷(N[1]-2))×((N[1]+1)×÷N[1])×(1-RS[8;5])
[12]  V[6]←V[6]+1-((N[1]-1)×÷(N[1]-3))×((N[1]+5)×÷N[1])×(1-RS[8;5])
[13]  CV1← 5 4 3 2 1 REG SAMP[(⍳(1+ρSAMP)-N2);]
[14]  V[7]←V[7]+CV1[8;5]
[15]  CV2←SAMP[⍳N2;1],[1.5] CV2←RS[1;2]+CV2←SAMP[⍳N2; 2 3 4 5]+.×RS[5 4 3 2 ;2]
[16]  CV2← 2 1 REG CV2
[17]  V[8]←V[8]+CV2[5;5]
[18]  →GO×⍳N[2]>I←I+1
[19]  V←V÷N[2]
[20]  RS← 5 4 3 2 1 REG POP
[21]  'THE MULTIPLE R SQ FOR THE POPULATION IS ', 6 4 ▼RS[8;5]
[22]  'THE SAMPLE SIZE = ',(5 0 ▼N[1]),'          SAMPLES TAKEN = ', 5 0 ▼N[2]
[23]  'THE AVERAGE SHRINKAGE ESTIMATES FOR THESE SAMPLES ARE:'
[24]  '        RS      RS(W)      RS(M)      RS(L1)     RS(L2)      RS(D)      CV(1)     *CV(2)*'
[25]  10 4 ▼V
      ∇
```

20

```
      ∇CM[□]∇
      ∇ R←CM X;V
[1]   R←R÷(V∘.×V←(1 1)⍴R←(⍴R)+.×R←X-(ρX)ρ(+⌿X)÷(ρX)[1])*0.5
      ∇
```

21

```
      ∇RAND1[□]∇
      ∇ R←RAND1 N;K
[1]   K←0
[2]   R←,⍳0
[3]   LOOP:R←R,?500
[4]   →LOOP×⍳N>K←K+1
      ∇
```